Fermi National Accelerator Laboratory

FERMILAB-Conf-93/321-A
October 1993

COSMOLOGICAL PHASE TRANSITIONS

Edward W. KOLB
Fermi National Accelerator Laboratory and *The University of Chicago*

Abstract

If modern ideas about the role of spontaneous symmetry breaking in fundamental physics are correct, then the Universe should have undergone a series of phase transitions early in its history. The study of cosmological phase transitions has become an important aspect of early-Universe cosmology. In this lecture I review some very recent work on three aspects of phase transitions: the electroweak transition, texture, and axions.

1. Introduction

Perhaps the most important concept in modern particle theory is that of spontaneous symmetry breaking (SSB). The idea that there are underlying symmetries of Nature that are not manifest in the structure of the vacuum appears to play a crucial role in the unification of the forces. In all unified gauge theories—including the standard electroweak model—the underlying gauge symmetry is larger than the unbroken $SU(3)_C \otimes U(1)_{EM}$. Of particular interest for cosmology is the theoretical expectation that at high temperatures, symmetries that are spontaneously broken today were restored, and that during the evolution of the Universe there were phase transitions associated with spontaneous breakdown of gauge (and perhaps global) symmetries. For example, we can be reasonably confident that there was such a phase transistion at a temperature of order 300 GeV and a time of order 10^{-11} sec, associated with the breakdown of $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{EM}$. Moreover, the vacuum structure in many spontaneously broken gauge theories is very rich: Topologically stable configurations of gauge and Higgs fields exist as domain walls, cosmic strings, and monopoles. In addition, classical configurations that are not topologically stable, so-called nontopological solitons, may exist and be stable for dynamical reasons. Interesting examples include soliton stars, Q-balls, nontopological cosmic strings, global texture, sphalerons, and so on.

2. The Electroweak Phase Transition

The possibility that the baryon number of the Universe can be generated at the electroweak phase transition has triggered a lot of interest in understanding the dynamics of

To appear in the *Proceedings of the 37th Yamada Conference*, June 8-12, 1993, Tokyo, Japan

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

weakly first-order phase transitions in the early Universe. Since the original work of Kuzmin, Rubakov, and Shaposhnikov (1985) a great deal of effort has been dedicated to the construction of viable models that could generate the required baryon asymmetry.

As is well-known, one of the necessary ingredients of a successful baryogenesis scenario is a departure from equilibrium. Most scenarios of electroweak baryogenesis rely on the first-order nature of the phase transition to generate the required out-of-equilibrium conditions in the decay of the symmetric metastable phase by the nucleation of bubbles of the broken-symmetric phase. Baryon number is generated by the expansion of the bubble wall either by the scattering of heavy fermions off the wall, or by the unwinding of topologically non-trivial configurations in its neighborhood [for a review, see Cohen *et al.*., 1993].

Although it is now generally believed that a successful baryogenesis scenario at the electroweak scale requires a departure from the minimal electroweak model, understanding the dynamics of the electroweak phase transition is a crucial ingredient for any viable scenario. One of the main obstacles to a comprehensive study of the electroweak phase transition is our lack of knowledge of the correct effective potential that describes the system in the vicinity of the critical temperature, T_C. Problems due to infrared divergences have been known since the original work of Dolan and Jackiw (1974) and Weinberg (1974). For the current limits on the masses of the Higgs and top quark, the 1-loop effective potential predicts a weak first-order transition. This is somewhat unsettling, because we know that weak first-order transitions have large infrared divergences which are not accounted for by the 1-loop calculation. In other words, if the 1-loop potential predicts a weak first-order transition, chances are that the actual transition is even weaker, if not actually second order. It is thus important to incorporate the infrared corrections to the effective potential. In fact, a few recent works have incorporated some infrared corrections caused by the vanishing of vector boson masses near $\phi = 0$ by summing over ring, or daisy, diagrams. As clearly shown in the paper by Dine *et al.* (1992), these corrections decrease the effective tunneling barrier for decay, weakening the strength of the transition. The validity of the ring-improved effective potential for the temperatures of interest relies on cutting off higher-order contributions by invoking a non-perturbative magnetic plasma mass, $M_{\rm plasma}$, for the gauge bosons such that the loop expansion parameter, $g^2T/M_{\rm plasma}$, is less than 1. Since this non-perturbative contribution is not well understood at present, one should take the results from the ring-improved potentials with some caution.

In addition to infrared problems caused by the vanishing of the vector boson masses in the symmetric phase, for sufficiently weak transitions we point out that one must take into account infrared divergences caused by the small Higgs mass if $M_H \ll T$. We will review the well known formalism for field theory at high temperatures. We point out that the loop expansion parameter diverges as the Higgs mass vanishes. This means that diagrams contributing to the 1-loop potential that are unimportant at either high or low temperatures may be dominant around the critical temperature. For instance, at zero temperature the loop expansion parameter for the Higgs loops is λ, the Higgs self coupling. However for the 3-dimensional effective field theory at high temperature, the loop expansion parameter is $\lambda T/M_H(T)$ for the Higgs loops. For $T \gg M_H(T)$, the loop expansion is not under control. We point out that this is exactly the situation in the standard electroweak model between the critical temperature and the spinodal temperature where $M_H(T)$ vanishes.

In this talk I will attempt to estimate the magnitude of the infrared corrections to the 1-loop electroweak potential near the critical point using a familiar technique from condensed-matter physics. In particular, I will argue that near the critical point it is possible to estimate the fluctuations in the spatial correlations of the magnitude of the scalar field by using well known results from the theory of critical phenomena. I will show that the effective corrections to the critical exponent that controls the behavior of the correlation length for the electroweak model can be approximated by considering an associated Ginzburg–Landau (G–L) model just above its critical temperature. This approach has been successfully implemented by De Gennes (1973) in the study of liquid crystals, and recently by Fernández *et al.* (1992) in the study

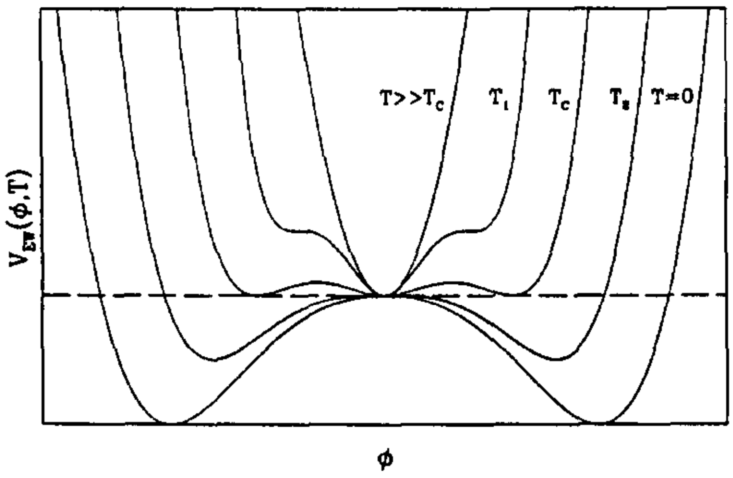

Fig. 1. One-loop electroweak potential at several different temperatures.

of the 7-states Potts model which exhibits a weak first-order transition. It is instructive to examine the critical behavior of the 1-loop effective potential for the electroweak model.

The 1-loop finite-temperature corrections to the electroweak potential have been studied in detail in the literature, most recently by Anderson and Hall (1992). They showed that a high temperature expansion of the 1-loop potential closely approximates the full 1-loop potential for $M_H \lesssim 150$ GeV and $M_T \lesssim 200$ GeV. (It is important to differentiate between the finite temperature Higgs mass, $M_H(T)$ and the zero-temperature Higgs mass, M_H.) They obtain for the potential

$$V_{\rm EW}(\phi,T) = D\left(T^2 - T_2^2\right)\phi^2 - ET\phi^3 + \frac{1}{4}\lambda_T\phi^4, \tag{1}$$

where D and E are given by $D = \left[6(M_W/\sigma)^2 + 3(M_Z/\sigma)^2 + 6(M_T/\sigma)^2\right]/24$, and $E = \left[6(M_W/\sigma)^3 + 3(M_Z/\sigma)^3\right]/12\pi$. Here T_2 is the temperature at which the origin becomes an inflection point (*i.e.*, below T_2 the symmetric phase is unstable and the field can classically evolve to the asymmetric phase by the mechanism of spinodal decomposition), and is given by

$$T_2 = \sqrt{(M_H^2 - 8B\sigma^2)/4D}\ . \tag{2}$$

The physical Higgs mass is given in terms of the 1-loop corrected λ as $M_H^2 = (2\lambda + 12B)\sigma^2$, with $B = \left(6M_W^4 + 3M_Z^4 - 12M_T^4\right)/64\pi^2\sigma^4$. We use $M_W = 80.6$ GeV, $M_Z = 91.2$ GeV, and $\sigma = 246$ GeV. The temperature-corrected Higgs self-coupling is

$$\lambda_T = \lambda - \frac{1}{16\pi^2}\left[\sum_B g_B\left(\frac{M_B}{\sigma}\right)^4 \ln\left(M_B^2/c_BT^2\right) - \sum_F g_F\left(\frac{M_F}{\sigma}\right)^4 \ln\left(M_F^2/c_FT^2\right)\right] \tag{3}$$

where the sum is performed over bosons and fermions (in our case only the top quark) with their respective degrees of freedom $g_{B(F)}$, and $\ln c_B = 5.41$ and $\ln c_F = 2.64$.

Apart from T_2, there will be two temperatures of interest in the study of the phase transition. For high temperatures, the system will be in the symmetric phase with the potential exhibiting only one minimum at $\langle\phi\rangle = 0$. As the Universe expands and cools an inflection point will develop away from the origin at

$$\phi_{\rm inf} = 3ET_1/2\lambda_T, \tag{4}$$

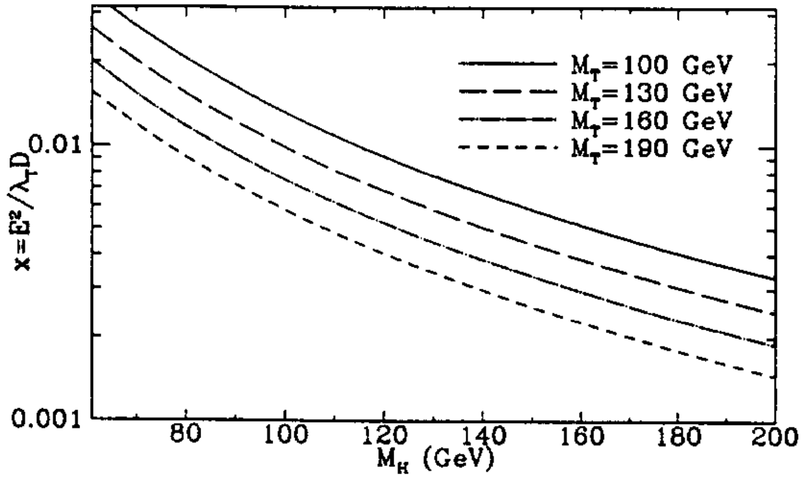

Fig. 2. Parameter $x = E^2/\lambda_T D$ as a function of the Higgs boson mass for several values of the top quark mass.

where T_1 is given by

$$T_1 = T_2/\sqrt{1 - 9E^2/8\lambda_T D}\,. \tag{5}$$

For $T < T_1$, the inflection point separates into a local maximum at ϕ_- and a local minimum at ϕ_+, with $\phi_\pm = \{3ET \pm [9E^2T^2 - 8\lambda_T D(T^2 - T_2^2)]^{1/2}\}/2\lambda_T$. At the critical temperature

$$T_C = T_2/\sqrt{1 - E^2/\lambda_T D}\,, \tag{6}$$

the minima have the same free energy, $V_{\rm EW}(\phi_+, T_C) = V_{\rm EW}(0, T_C)$. (Note that $V(\phi,T)$ is the homogeneous part of the free energy density whose minima denote the equilibrium states of the system. Accordingly, in this work we freely interchange between calling $V(\phi,T)$ a potential and a free energy density.)

In Fig. 1 we show the electroweak potential at temperatures $T \gg T_1$, T_1, T_C, T_2, and $T = 0$. The difference between the temperatures T_1, T_C, and T_2 is determined by the parameter

$$x = E^2/\lambda_T D. \tag{7}$$

This parameter is shown in Fig. 2 for different values of M_H and M_T. Clearly $x \ll 1$ for the minimal electroweak model, so we can write the approximate relations

$$\begin{aligned} T_C &\simeq T_2(1 + x/2) \\ T_1 &\simeq T_2(1 + 9x/16). \end{aligned} \tag{8}$$

It is useful to understand why the transition is first order; *i.e.*, why at T_C there is a barrier between the high-temperature phase and the low-temperature phase. It has been appreciated for a long time that a pure $\lambda\phi^4$ theory is equivalent to a Ginzburg–Landau theory, which has a second-order phase transition. The reason the electroweak theory is first order, rather than second order, is that there is an additional attractive force between scalar particles mediated by the vector bosons. This additional attractive force leads to a condensate of the Higgs field at a temperature slightly above T_2. T_2 and T_C would be the same (a second-order transition) in the absence of gauge boson interactions. (Note that as $E \to 0$, *i.e.*, as vector interactions are turned off, $T_C \to T_2$.)

The whole picture of bubble nucleation relies on the behavior of $V_{\rm EW}(\phi,T)$ between T_C and T_2. In the standard picture, one assumes that the system is in a near-homogeneous state around its equilibrium value (in this case $\langle\phi\rangle = 0$), so that large thermal fluctuations in the spatial correlations of ϕ are exponentially suppressed above the scale of the thermal correlation length, $\xi(T)$,

$$\xi^{-2}(T) \equiv M_H^2(T) = \frac{\partial^2 V_{\rm EW}(\phi = \langle\phi\rangle, T)}{\partial\phi^2}. \tag{9}$$

In this case, for some temperature $T_C > T > T_2$, critical bubbles of the broken-symmetric phase appear and expand. They eventually collide with other bubbles, converting the symmetric phase into the broken-symmetric phase.

For the electroweak potential the difference between T_C and T_2 is very small: $\eta_2(T_C) \equiv (T_C - T_2)/(T_C + T_2) \sim x/4 \ll 1$. The transition is predicted to be weakly first order. As mentioned above, infrared corrections to the 1-loop potential can be very important due to its flatness (small mass) around $\phi = 0$. As we shall see below, the loop expansion parameter for the Higgs loops at high temperatures is not λ, but $\lambda_T T/M_H(T)$. We can estimate where this will become large for the standard electroweak model. Before starting, it is helpful to note that the temperature-corrected Higgs self-coupling, λ_T, is approximately equal to the *tree-level* Higgs self-coupling, $\lambda_0 = M_H^2/2\sigma^2$. For the electroweak potential near T_C, $M_H^2(T_C) = 2D(T_C^2 - T_2^2)$. Since $T_2^2/T_C^2 = 1 - E^2/\lambda_T D$, $M_H(T_C) = T_C E\sqrt{2/\lambda_T}$. Therefore at T_C the loop expansion parameter is $\lambda_T T_C/M_H(T_C) = \lambda_T^{3/2}/E\sqrt{2}$. Now as discussed above, to a reasonable accuracy $\lambda_T = M_H^2/2\sigma^2$ (here, of course, M_H is the zero-temperature mass). Thus

$$\lambda_T T_C/M_H(T_C) \simeq M_H^3/4E\sigma^3 \sim 1.68(M_H/100 \text{ GeV})^3. \tag{10}$$

For M_H greater than about 84 GeV, at T_C the expansion parameter exceeds unity. Between T_C and T_2 the mass goes to zero, so the corrections are even larger.

The question we would like to address is, can we estimate the magnitude of the infrared corrections in a simple way? Since we are interested in the behavior of the system around $\langle\phi\rangle = 0$ for $T_C \geq T \geq T_2$, we will show that it is possible to map the electroweak potential *in a small neighborhood around* $\phi = 0$ to an effective Ginzburg–Landau (G–L) theory which exhibits a second-order phase transition at T_2. The critical behavior of this model has been extensively studied in the seventies using renormalization group (RG) techniques pioneered by Wilson. In particular, infrared corrections to the G–L model which are important around the critical temperature have been computed using ε-expansion techniques. The net result is that the magnitude of fluctuations on the spatial correlations of the order parameter calculated by mean-field theory (which we will show is equivalent to the 1-loop potential) is largely underestimated. We will obtain the corrections to the G–L model and map it back to the electroweak potential in an attempt to estimate the infrared corrections to the 1-loop result. We will show that the corrections to tunneling rates can be very large, indicating the failure of the naïve 1-loop potential to describe the dynamics of the transition.

2.1. Critical behavior of ϕ^4 field theory

In order to study the critical behavior of a ϕ^4 scalar field theory we follow Ginsparg (1980) in reducing the theory to an effective theory of the static mode of the scalar field in $d = 3$ dimensions. The generating functional in the presence of a source $J(x)$ for a zero temperature scalar field theory in Euclidean ($t = -i\tau$) space-time is (we use $\hbar = c = 1$)

$$Z[J] = \int [\mathcal{D}\phi] \exp\left\{-\int d^4x \left[\frac{1}{2}(\partial\phi)^2 - \frac{1}{2}m^2\phi^2 + \frac{\lambda}{4}\phi^4\right] - \int d^4x J\phi\right\}. \tag{11}$$

In order to study the theory at finite temperature we take the Euclidean time to be periodic in β, and sum only over periodic paths with $\phi(0,\mathbf{x}) = \phi(\tau,\mathbf{x})$, as is well known. Due to the periodic behavior in τ we can expand the scalar field as

$$\phi(\tau,\mathbf{x}) = \frac{1}{\beta}\sum_{n=-\infty}^{+\infty}\int\frac{d^3k}{(2\pi)^3}\exp\left(i\omega_n\tau + i\mathbf{k}\cdot\mathbf{x}\right)\phi_n(\mathbf{k}); \qquad \omega_n = 2\pi n/\beta. \tag{12}$$

By rescaling the field $\phi_n(\mathbf{k})$ by $\beta^{-1/2}$, and separating the static ($n = 0$) mode from the rest, we obtain,

$$\begin{aligned} Z[J] &= \int_{\text{periodic}}[\mathcal{D}\phi]\exp\Big\{-\int_{\mathbf{k}}\frac{1}{2}\left(\mathbf{k}^2 - m^2\right)\phi_0(\mathbf{k})\phi_0(-\mathbf{k}) \\ &\quad -\sum_{n\neq 0}\int_{\mathbf{k}}\frac{1}{2}\left[(2\pi n/\beta)^2 + \mathbf{k}^2 - m^2\right]\phi_n(\mathbf{k})\phi_{-n}(-\mathbf{k}) + \int J\phi \\ &\quad -\frac{\lambda/\beta}{4}\sum_{n,n',n''=-\infty}^{+\infty}\int_{\mathbf{k},\mathbf{k}',\mathbf{k}''}\phi_n(\mathbf{k})\phi_{n'}(\mathbf{k}')\phi_{n''}(\mathbf{k}'')\phi_{-n-n'-n''}(-\mathbf{k}-\mathbf{k}'-\mathbf{k}'')\Big\} \end{aligned} \tag{13}$$

where $\int_{\mathbf{k}} = \int d^3k/(2\pi)^3$. The effective $d = 3$ theory is obtained by summing over all the $n \neq 0$ modes. Perturbatively, this means that all internal lines in the Feynman diagrams will correspond to sums over the $n \neq 0$ modes, and the external lines are given only by the $n = 0$ mode. This way the higher modes will contribute to the mass, wave-function renormalizations, and to the N-point function for the effective theory of the $n = 0$ mode. It is then possible to construct an effective Lagrangian $\mathcal{L}_{\text{eff}}$ for the $d = 3$ theory by a systematic perturbation expansion in λ. The leading contribution to the 2-point function is given by the tadpole diagram obtained by summing over the higher modes in the Feynman propagator. One obtains to leading order,

$$\mathcal{L}_{\text{eff}} = \frac{1}{2}\left(-m^2 + \frac{\lambda}{4\beta^2} + \ldots\right)\phi_0(\mathbf{k})\phi_0(-\mathbf{k}) + \frac{\lambda/\beta}{4}\phi_0^4 + \ldots. \tag{14}$$

Note that in the effective $d = 3$ theory the coupling λ is dimensionful with the temperature naturally setting its dimensionality. This temperature dependence will not be relevant for the discussion of the critical behavior of the theory and will be absorbed into the definition of λ. The static critical behavior of the original $d = 4$ theory is completely embodied in the effective theory for the $n = 0$ mode in $d = 3$. To leading order in λ, this theory exhibits a phase transition at a critical temperature

$$T_C^2 = 1/\beta_C^2 = \frac{4m^2}{\lambda}\left(1 + O(\lambda) + \ldots\right). \tag{15}$$

It turns out that the critical behavior of the theory above, known as the G–L model, has been extensively studied using the ε-expansion to incorporate higher-order infrared effects. From $\mathcal{L}_{\text{eff}}$ we can obtain the effective 3-dimensional potential to leading order in λ,

$$V_{\text{G-L}}(\phi,T) = \frac{m^2(T)}{2}\phi^2 + \frac{\lambda}{4}\phi^4; \qquad m^2(T) \equiv \frac{\lambda}{4}\left(T^2 - T_C^2\right), \tag{16}$$

where $\phi(\mathbf{x})$ is the static scalar field, which is the relevant order parameter in equilibrium.

As is well-known, this theory exhibits a second-order phase transition at T_C; above T_C the left-right symmetry is exact and the equilibrium value of ϕ is $\langle\phi\rangle$=0. Below T_C the symmetry is broken and the equilibrium value of ϕ is $\langle\phi\rangle = \pm\left[m^2(T)/\lambda\right]^{1/2}$. In the thermodynamic limit, the system will eventually settle at one value of ϕ, since any interface is energetically unfavored. Of course $\langle\phi\rangle$ only gives information about the homogeneous behavior of ϕ. Typically, there will be fluctuations around $\langle\phi\rangle$ which are correlated within the correlation length scale defined in Eq. (9). For temperatures above and below T_C (denoted by $+$ and $-$ respectively) we obtain from Eq. (16)

$$\xi_+^{-2}(T) = m^2(T) = \frac{\lambda}{4} T_C^2 (1 + T/T_C)^2 \left(\frac{T - T_C}{T + T_C}\right), \tag{17}$$

and

$$\xi_-^{-2}(T) = -2m^2(T) = \frac{\lambda}{2} T_C^2 (1 + T/T_C)^2 \left(\frac{|T - T_C|}{T + T_C}\right). \tag{18}$$

This is the well known result from mean-field theory, usually expressed as

$$\xi_{\mathrm{MF}}(T) \propto |T - T_C|^{-\nu}; \qquad \nu = 1/2, \tag{19}$$

where the critical exponent ν expresses the singular behavior of $\xi(T)$ as $T \to T_C$ both from above and below. It is clear that fluctuations around the equilibrium value of ϕ are indeed very large near T_C, being considerably larger than the mean field results. Thus, the assumption of near homogeneity is not valid if T is sufficiently close to T_C.

In order to handle the infrared divergences that appear near T_C, the RG is used to relate a given theory to an equivalent theory with larger masses and thus better behaved in the infrared. Within the ε expansion, one works in $4 - \varepsilon$ dimensions and finds a fixed point of order ε of the RG equations, taking the limit $\varepsilon \to 1$ in the end. To second-order in ε one obtains,

$$\nu = \frac{1}{2} + \frac{1}{12}\varepsilon + \frac{7}{162}\varepsilon^2 \simeq 0.63. \tag{20}$$

The corrected critical exponent embodies corrections coming from the infrared divergences near T_C. The ε-corrected correlation length can be written above T_C as

$$\left[\xi_+^{\varepsilon}(T)\right]^{-1} = \sqrt{\frac{\lambda}{4}}\, T_C \,(1 + T/T_C) \left(\frac{T - T_C}{T + T_C}\right)^{0.63}. \tag{21}$$

Below T_C we obtain,

$$\left[\xi_-^{\varepsilon}(T)\right]^{-1} = \sqrt{\frac{\lambda}{2}}\, T_C \,(1 + T/T_C) \left(\frac{|T - T_C|}{T + T_C}\right)^{0.63}, \tag{22}$$

so that, in both cases the ratio between the mean field and ε-corrected correlation lengths can be written as

$$\frac{\xi_{\mathrm{MF}}(T)}{\xi_\epsilon(T)} = \eta_C^{0.13}(T); \qquad \eta_C(T) \equiv \frac{|T - T_C|}{T + T_C}. \tag{23}$$

If we are interested in studying the behavior of the theory above T_C we can use the fact that $\xi(T) = m^{-1}(T)$ to obtain an ε-corrected mass,

$$m_\epsilon(T) = \eta_C^{0.13}(T) m(T). \tag{24}$$

A similar result can be easily obtained below T_C.

2.2. *Infrared corrections to the electroweak potential*

Now I will argue that we can obtain information on the critical behavior of the electroweak phase transition between T_C and T_2 by studying a G–L model with a critical temperature that we take to be T_2. This is possible since T_C is so close to T_2 due to the weakness of the transition already at 1-loop level. [See Fig. 2.] Thus, we will estimate the infrared corrections to the electroweak model by looking at the G–L model around T_C. Clearly this is only an approximation to treating the full problem of incorporating the ε-expansion for the

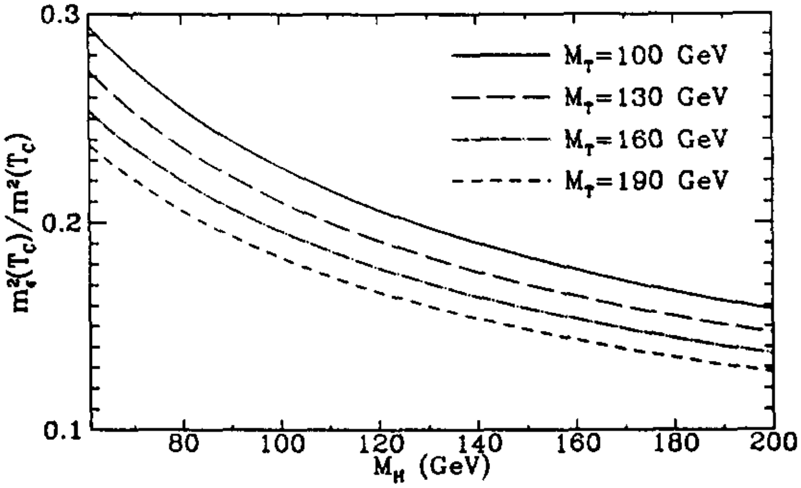

Fig. 3. ϵ-corrected mass as a function of the Higgs mass for several values of the top mass.

standard model. However, from the nature of the potential, we claim that our results are a lower bound on the true infrared corrections, which we conjecture will be even more severe than what we will estimate below.

Start with the simplest possible approach, by studying the G–L model defined by the free energy density,

$$V_{\rm G-L}(\phi,T) = \frac{m^2(T)}{2}\phi^2 + \frac{\lambda_T}{4}\phi^4; \qquad m^2(T) \equiv 2D\left(T^2 - T_2^2\right), \tag{25}$$

where D, T_2, and λ_T are defined as above. This is simply $V_{\rm EW}(\phi,T)$ with $E \to 0$. This model exhibits a second-order phase transition at $T = T_2$. Recall that this is the temperature at which the barrier disappears in the 1-loop electroweak potential. [See Fig. 1.] Thus, we are interested in the behavior of this model for temperatures above T_2. The claim is that for $T \lesssim T_C$ and in the neighborhood of $\langle\phi\rangle = 0$ this model can be used to give us an *estimate* of the infrared corrections to the electroweak potential. Note that our choice of the mass is such that the correlation length for fluctuations around equilibrium is the same in both models. Thus, the behavior around $\langle\phi\rangle = 0$ is well-matched by the G–L model

From the results of the previous section, the ε-corrected mass is

$$m_\varepsilon^2(T) = 2D\eta_2^{0.26}(T)\left(T^2 - T_2^2\right); \qquad \eta_2(T) = \frac{|T - T_2|}{T + T_2}. \tag{26}$$

The value of $\eta_2(T)$ at $T = T_C$ can be found using T_C and T_2 from Eqs. (6) and (2):

$$\eta_2(T_C) = \frac{1 - \sqrt{1 - E^2/\lambda_T D}}{1 + \sqrt{1 - E^2/\lambda_T D}}. \tag{27}$$

In Fig. 3 we show $m_\varepsilon^2(T_C)/m^2(T_C) = \eta_2^{0.26}(T_C)$ as a function of the Higgs mass for several values of the top mass. It is clear that the infrared corrections are quite large for all values of parameters probed. Below T_C the potential is even flatter near the origin and the infrared problem is even more severe. For larger values of ϕ the cubic term becomes important increasing the flatness of the electroweak model compared to the G–L model (leading again to more severe infrared problems). Again we stress that this is not intended to be an exact calculation of the infrared corrections to the electroweak potential, but simply an estimate of the magnitude of these corrections for small ϕ. As mentioned earlier, we expect the true corrections to be even more severe that what we obtained above.

As a possible application of the above results, we estimate the corrections to the 1-loop tunneling rate using $m_\varepsilon^2(T)$. This is clearly an approximation since we have stressed that our approach is only valid in a small neighborhood of $\langle\phi\rangle = 0$, and should not be trusted for $\phi \gtrsim D(T^2 - T_2^2)/ET$, for a given T. We want to estimate how severe the corrections to tunneling could be due to the smallness of the curvature at the origin. The finite-temperature tunneling rate, $\Gamma \propto \exp(-S_3/T)$, for a theory with a potential like the electroweak potential has been shown by Dine *et al.* to have an approximate analytical expression for the exponent given by

$$\frac{S_3}{T} = 4.85\frac{m^3(T)}{E^2T^3}f(\alpha) \qquad \alpha = \frac{\lambda_T m^2(T)}{2E^2T^2}, \tag{28}$$

with

$$f(\alpha) = 1 + \frac{\alpha}{4}\left[1 + \frac{2.4}{1-\alpha} + \frac{0.26}{(1-\alpha)^2}\right]. \tag{29}$$

However, according to our arguments, for $T < T_C$ the effective curvature of the potential around the equilibrium point is smaller than what is estimated from the 1-loop approximation. The effective tunneling barrier is then also smaller, and the kinetics of the transition may be different from the usual nucleation scenario. [An interesting possibility is that the critical temperature for the corrected theory is larger than the 1-loop result. However, as we remarked earlier, our method is only applicable in a small neighborhood of ϕ, and we cannot use it to study the potential away from the origin which is necessary to predict T_C in a first order phase transition.] Taking into account the ε-corrections above, the exponent becomes,

$$\frac{S_3^\varepsilon}{T} = 4.85\eta_2^{0.39}(T)\,\frac{m^3(T)}{E^2T^3}f\left(\alpha_\varepsilon\right); \qquad \alpha_\varepsilon = \eta_2^{0.26}(T)\,\frac{\lambda_T m^2(T)}{2E^2T^2}. \tag{30}$$

Clearly use of the ε-expansion improved mass can have an enormous effect upon the tunnelling rate, changing the *exponent* by a large factor.

2.3. *Thermal Fluctuations and Sub-Critical Bubbles*

Now I discuss how it is possible to examine the strength of a first-order transition by a simple method based on the work of Gleiser, Kolb, and Watkins (1991) estimating the thermal nucleation rate of "sub-critical bubbles," which are correlation volume fluctuations of one phase inside the other phase. We will argue that when applied to the electroweak phase transition, it gives results which are qualitatively consistent, signaling the failure of the naïve 1-loop potential as a valid approximation to study the dynamics of the transition. Since we have recently applied this method to the electroweak potential of Eq. (1), we will be quite brief here and refer the reader to Gleiser, Kolb, and Watkins for details.

Consider the electroweak potential of Fig. 1. Below T_1 a new minimum develops at ϕ_+ away from the symmetric minimum at $\langle\phi\rangle = 0$. There will be a non-zero probability for bubbles of radius R of the new phase at ϕ_+ to be thermally nucleated. The thermal nucleation rate for producing a bubble of radius R is given by $\Gamma(R,T) \sim \exp[-F(R)/T]$, where $F(R)$ is the free energy of the fluctuating region of radius R. For $T \geq T_C$ it is clear that the larger the bubble the more unfavored it is, since the free energy is a monotonically increasing function of R. The bubbles will shrink in a time scale determined by many factors. For example, for a curvature dominated motion of the bubble wall, which is probably a good approximation close to T_C, the radius of a large bubble shrinks as $t^{1/3}$. Some recent numerical studies showed that even small bubbles persist longer than one would naïvely estimate, bouncing back a few times before dissipating all their energy into quanta of the field. However, due to the exponential suppression in their production rate, unless the transition is very weakly first order (with the whole bubble picture being invalid in this case), only bubbles with small enough radius can be efficiently produced so that at any given time a reasonable fraction of the horizon volume can be occupied by the new phase at ϕ_+. Although there is a distribution of bubbles with different radii, it is clear from the above arguments that bubbles with a correlation volume will be statistically dominant. [The kinetics of the transition is bound to be much more complicated than these simple arguments may imply. There will be many different processes contributing to the number density of bubbles of a given radius as a function of time, such as capture and evaporation of particles from bubbles, coalescence due to bubble collisions, shrinking of larger bubbles, neighbor-induced nucleation, and possible shape instabilities, to name just a few.]

The basic idea behind the sub-critical bubbles method is that for sufficiently weak first order transitions, the rate for producing bubbles of a correlation volume is quite large, so that at any given time there will be an appreciable fraction of the total volume occupied by the new phase. If this is the case, the usual assumption of near-homogeneity used in vacuum decay calculations is not valid; instead of having critical bubbles being nucleated on a background of the metastable phase, nucleation would occur in a background which is better described by a dilute gas of small, non-perturbative fluctuations. There is no reason to expect that the usual calculation for the decay rate is applicable in this case.

The free energy of a spherically symmetric fluctuation around equilibrium is

$$F(T) = 4\pi \int_0^\infty r^2\, dr \left[\frac{1}{2}\left(\frac{d\phi}{dr}\right)^2 + V_{\mathrm{EW}}(\phi, T)\right]. \tag{31}$$

We will focus on the electroweak model at T_C. In principle, there will be fluctuations from $\phi = 0$ to ϕ_+ and back, although at T_C the free energies for these fluctuations are identical. The rates for the thermal fluctuations can be estimated by making an ansatz for the radial profile of the sub-critical bubbles. Following Gleiser, Kolb, and Watkins, we write

$$\phi_+(r) = \phi_+ \exp\left(-r^2/\ell^2\right), \tag{32}$$

where $\phi_+(r)$ corresponds to a bubble of broken phase ϕ_+ nucleated in the symmetric phase $\phi = 0$. The parameter ℓ controls the approximate size of the bubble which we take to be the correlation length. Introducing the dimensionless variables $X(\rho) = \phi(r)/\sigma$, $\tilde{\ell}(T) = \ell(T)\sigma$, $\theta = T/\sigma$, and $\rho = r\sigma$, we obtain

$$F_+(\theta) = \pi^{3/2} X_+^2 \tilde{\ell}\sigma \left[\frac{3\sqrt{2}}{8} + \tilde{\ell}^2\left(\frac{D\sqrt{2}}{4}\left(\theta_C^2 - \theta_2^2\right) - \frac{E\theta_C\sqrt{3}}{9} X_+ + \frac{\lambda_T}{32} X_+^2\right)\right]. \tag{33}$$

In order for sub-critical bubbles to be of cosmological relevance, their thermal nucleation rate must be considerably larger than the expansion rate of the Universe, $\Gamma(\xi(T), T)/H > 1$, with $H \simeq 1.66 g_*^{1/2} T^2/M_{\mathrm{PL}}$, where $g_* \simeq 110$ is the number of effective relativistic degrees of freedom at the electroweak scale. Neglecting pre-factors, this condition can be easily seen to lead to the inequality $F_+(T)/T < 34$. It is clear that at T_C this condition is comfortably satisfied for the present lower bound on the Higgs mass, $M_H \geq 57$ GeV, for which we obtain $\Gamma(\xi, T_C)/H \sim 10^8$.

Recently, Dine *et al.* argued that sub-critical bubbles would not be of relevance for most (if not all) the parameter space of the standard model due to the smallness of the thermal dispersions around $\langle\phi\rangle = 0$. We agree with their results for $M_H \sim 60$ GeV. However, for larger Higgs masses fluctuations in long wavelengths are quite large, contrary to their claim. We hoped to have shown here that both the estimate from the thermal dispersion and from sub-critical bubbles indicate that there will be large fluctuations around equilibrium, signaling the failure of the 1-loop potential to describe the dynamics of the transition.

2.4. *Electroweak Outlook*

In this work we have argued that it is possible to study the critical behavior of a weak first order transition which has a spinodal instability at some temperature T_2 by mapping its behavior around equilibrium, $\langle\phi\rangle$, to an effective Ginzburg-Landau model above its critical temperature T_2. In this way, both models have the same spinodal instability at $\langle\phi\rangle$ so that infrared corrections can be estimated from well-known ε-expansion methods. This approach is completely general and can in principle be applied to any sufficiently weak first order transition. It suits the standard electroweak model particularly well due to the closeness of its critical temperature T_C to the spinodal instability temperature T_2. In fact, the difference between the two temperatures should provide a qualitative measure of the weakness of the transition.

Incorporating the ε-expansion results leads to a larger correlation in the spatial fluctuations of the order parameter, which can be translated into a smaller (infrared corrected) mass for excitations around $\langle\phi\rangle$. Thus, the strength of the transition is considerably weaker than one would estimate from the naïve 1-loop potential. We do not claim here to have obtained the ε-corrected effective potential, but an estimate of the infrared corrections which are not included in the 1-loop result. Our results provide a simple way to examine the importance of these corrections around T_C, offering a simple way of estimating the strength of the transition. If the η parameter is close to unity at the critical temperature T_C the transition is well described by the 1-loop result. Otherwise, the transition is weakly first order, and one should be very careful when adopting the usual vacuum decay formalism to study the transition. For temperatures just below or above T_C, the large fluctuations around equilibrium are quite apparent. In particular, one can picture the behavior of this system just below its critical temperature as being qualitatively similar to the behavior of a weak first-order transition *at* its critical temperature.

The work described here is in some sense a different approach to the problem of the infrared problems in the perturbation expansion studied by many people. In most works, heroic efforts are made to isolate and calculate the most important contributions of the higher order diagrams. Here we are simply able to estimate the magnitude of the corrections. They are obviously going to be quite large, and any detailed model for electroweak baryogenesis will probably have to wait until these problems are solved.

For more details on the work reported here, see Gleiser and Kolb (1993). This work also contains references to the important perturbative approach.

3. Gravity and Global Symmetries

3.1. *Texture*

For the most part, cosmologists have concentrated on the analysis of topological defects that can arise in *gauge* theories. However defects can also arise in the spontaneous breaking of *global* symmetries. The analogies of the local strings and monopoles are global strings and global monopoles. The global field configurations look like their local counterparts for the scalar field, but of course there is no vector field. This means that formally the string and monopole solutions have infinite energy (recall for the local defects the energy in the gauge fields cancels the energy in the Higgs field far from the defect.) This is really not a problem, because there the divergence in the energy is only logarithmic, and there are many physical effects to cut it off (such as the inter-defect separation). There are just two main differences in the behavior of gauge and global defects: 1) the energy of the global defects are slightly more spread out, 2) the global strings can radiate energy by the emission of Nambu–Goldstone bosons.

However there are new types of defects in global symmetry breaking that do not appear in the breaking of gauge symmetries. For example, in the spontaneous breaking of a global $O(N)$ model to $O(N-1)$, for $N=1$ walls appear, for $N=2$ global strings result, for

$N = 3$ global monopoles are produced. These all have counterparts in local theories. However for $N > 3$ global defects also exist: for $N = 4$ the defect is called global texture, and for $N > 4$ they are called Kibble gradients. Texture corresponds to knots in the Higgs field that arise when the field winds around the three sphere. These knots are generally formed by misalignment of the field on scales larger than the horizon at the symmetry breaking phase transition because of the Kibble mechanism. As the knots enter the horizon, they collapse at roughly the speed of light, giving rise to nearly spherical energy density perturbations. New knots are constantly coming into the horizon and collapsing, leading to a scale invariant spectrum of density perturbations. The magnitude of the perturbations is set by the scale of the symmetry breaking, and for scenarios of structure formation involving texture, the scale of symmetry breaking must be about 10^{16}GeV.

A theory of texture or Kibble gradients being responsible for the seeds of large-scale structure has been formulated by Turok, Spergel and collaborators (Pen *et al.*, 1993). Texture would provide a very promising alternative to conventional inflation scenarios for generating the primordial density fluctuations if indeed they are ubiquitous in particle physics models. In fact, texture arises in a variety of theories with nonabelian global symmetries that are spontaneously broken. However, even an extremely small amount of explicit symmetry breaking will spoil the texture scenario. I would like to close these lectures by discussing how sensitive this theory is to Planck-scale effects. This idea was recently discussed by Holman *et al.* (1992) and Kamionkowski *et al.* (1992).

To illustrate these possibilities, consider a theory with a global $O(N)$ symmetry spontaneously broken to $O(N-1)$ by an N-vector. The theory is described by the scalar potential $V(\Phi) = \lambda\left(\Phi^a\Phi^a - f_T^2\right)^2$. As mentioned above, texture arises for $N = 4$. There are many arguments suggesting that *all* global symmetries are violated at some level by gravity. For example, both wormholes and black holes can swallow global charge. "Virtual" black holes or wormholes, which should, in principle, arise in a theory of quantum gravity, will lead to higher dimension operators which violate the global symmetry. There are two possible assumptions one might make about the fate of global symmetries in a Universe that includes gravity. The *strong* assumption is that, despite all indications from low-energy, semi-classical gravitational physics (black holes, wormholes, etc.), it is possible to have exact global symmetries in the presence of gravity. This is the assumption made in the standard texture scenario. The *weak* assumption is that the global symmetry is not a feature of the full theory. There are two possible realizations of the weak assumption. Either the global symmetry is approximate, in which case one must include the effects of higher-dimensional, non-renormalizable, symmetry-breaking operators, or, consistent with indications from semi-classical quantum gravity, the global symmetry is never even an approximate symmetry unless protected by gauge symmetries.

If one makes the weak assumption, then one must include explicit symmetry breaking terms. If one assumes that gravity does not respect global symmetries at all, then renormalizable operators like ${M_{Pl}}^2\lambda_{ab}\Phi^a\Phi^b$, which explicitly break the global symmetry, should be included. These terms are expected, for instance, by the action of wormholes swallowing global charge. If virtual wormholes of size smaller than the Planck length are included, then we expect λ_{ab} to be of order unity. In this case it is wrong to consider an effective low-energy theory with a global symmetry. If one makes the assumption either that wormholes do not dominate the functional integral, or that the global charge is protected by gauge symmetries, then it may be possible to suppress the renormalizable operators. But even in this case higher dimension operators should be included. An example would be a dimension-5 operator, which would add to $V(\Phi)$ terms like $(\lambda_{abcde}/M_{Pl})\Phi^a\Phi^b\Phi^c\Phi^d\Phi^e$. Such terms explicitly break the global symmetry and lead to a mass for the pseudo-Nambu–Goldstone mode of $m^2 \propto \lambda f_T^3/M_{Pl}$. Of course the mass is suppressed by M_{Pl}, but we will show below that it still has a drastic effect on the texture scenario.

The implications of the strong and weak assumptions for texture are as follows: With the strong assumption, the texture scenario is unaffected. If one allows unsuppressed wormhole contributions, global symmetries (and hence texture) are a non-starter. If all effects of gravitational physics in the low-energy theory are contained in non-renormalizable terms, a more careful analysis is required. This is the possibility we explore now. In this approach we are then required to include all higher dimension operators consistent with the gauge symmetries of the model and suppressed by appropriate powers of M_{Pl}.

We now consider the effects of the higher dimension operators. These terms will break the symmetry explicitly, generating a complicated potential for the Nambu–Goldstone modes. In general, the vacuum manifold will be reduced to a point, though the potential will likely have many local minima. To see how this works, consider the theory discussed above with $N = 3$. Here, the vacuum manifold is the two sphere and the model, in two spatial dimensions, will have texture. (In three spatial dimensions, the model admits both global monopoles and texture, although the texture in this case is not spherically symmetric. We express the field as

$$\Phi = f\left(\sin\frac{\theta}{f_T}\cos\frac{\phi}{f_T},\ \sin\frac{\theta}{f_T}\sin\frac{\phi}{f_T},\ \cos\frac{\theta}{f_T}\right), \tag{34}$$

where θ and ϕ are the angular variables on the two-sphere which represent the Nambu–Goldstone modes of the problem.

The effect of the dimension 5 operators is to introduce 21 terms to the potential for the field which depend explicitly on θ and ϕ. (These are nothing more than the Y_{1m}, Y_{3m}, and Y_{5m} spherical harmonics.) Note that in general, the mass of the Nambu–Goldstone boson in this potential is roughly $f_T(f_T/M_{Pl})^{1/2}$.

So long as the mass of the Nambu–Goldstone mode is small compared to the Hubble parameter, the field will evolve essentially as in the original texture scenario. However, once the Compton wavelength of the Nambu–Goldstone mode enters the horizon, the field will begin to oscillate about the minimum (or rather the closest local minimum) of its potential. The field will then align itself on scales larger than the horizon and texture on all scales quickly disappear. For texture to be important for structure formation, they must persist at least until matter–radiation decoupling when $H \simeq 10^{-28}$eV.

The contribution of a dimension $4+d$ operator to the Nambu–Goldstone boson mass is $m \sim f_T(f_T/M_{Pl})^{d/2}$. Given that the texture scenario requires $f_T \sim 10^{16}$ GeV, the requirement that $m \lesssim 10^{-28}$eV implies that $d \gtrsim 35$; i.e., we must be able to suppress all operators up to dimension 40! It is rather difficult to see how this might occur; even the use of additional gauge quantum numbers could not prevent the occurrence of dimension 6 operators which break a non-Abelian symmetry (although they could protect an Abelian symmetry). We note that if we consider dimension-5 operators, then the mass becomes dynamically important immediately after the phase transition: texture therefore never exists.

3.2. Axions

It is well known that there are two contributions to CP violation in the standard model. First, QCD instantons induce a term $\mathcal{L}_{\rm QCD} = \theta\,{\rm tr}\,G\tilde{G}$ in the effective Lagrangian, which violates both P and CP. Here, θ is a dimensionless coupling constant, which one might naively expect to be of order unity. Second, the quark mass matrix can be complex, leading to a CP-violating phase in the Kobayashi-Maskawa mixing matrix. The phase of the quark mass matrix gives rise to an additional contribution $\theta_{\rm QFD} = \arg\det\mathcal{M}_{\rm q}$ to the coeffiecient of ${\rm tr}\,G\tilde{G}$. The degree of strong-CP violation is controlled by the parameter $\overline{\theta} = \theta + \arg\det\mathcal{M}_{\rm q}$, which is constrained by measurements of the electric dipole moment of the neutron to be less than 10^{-9}. The strong-CP problem is that there is no reason for these two contributions, which arise from entirely different sectors of the standard model, to sum to zero to such high accuracy.

The solutions that have been proposed for the strong-CP problem fall into three general classes. First, there are those that rely on the existence of an extra global $U(1)_A$ symmetry. This symmetry arises naturally if one or more of the quark masses are zero. In this case, it can be shown that the QCD θ parameter becomes unobservable. This solution is considered unattractive, since experimental evidence implies that it is unlikely that any of the quarks are massless. Peccei and Quinn (PQ) proposed a solution to the strong-CP problem in which they introduced an auxiliary, chiral $U(1)_{\rm PQ}$ symmetry that is spontaneously broken at a scale f_a, giving rise to a Nambu–Goldstone boson a known as the axion. This symmetry is explicitly broken by instanton effects. This explicit breaking generates a mass for the axion of order $m_a \sim \Lambda^2/f_a$, where Λ is the QCD scale. The important point is that the effective potential for the axion has its minimum at $\langle a/f_a\rangle = -\overline{\theta}$. It follows that when the axion field relaxes to its minimum, the coefficient of $\mathrm{tr}\, G\tilde{G}$ is driven to zero. This solution has received the most attention and has been explored by many authors.

A second class of solutions involve models where an otherwise exact CP is either softly or spontaneously broken. Specific models have been proposed where θ is calculably small and within the experimental limits.

A third class of solutions involve the action of wormholes. As we will argue below, wormholes can break global symmetries explicitly, thus giving rise to potentially large contributions to $\overline{\theta}$. However, under certain assumptions, it can be shown that wormholes actually have the effect of setting $\overline{\theta} = 0$.

Here, I address the question of whether these solutions to the strong-CP problem can remain viable if Planck scale effects break global symmetries explicitly. There are many arguments suggesting that all global symmetries are violated at some level by gravity. First, no-hair theorems tell us that black holes are able to swallow global charge. This allows for a gedanken experiment in which a quanta with global charge "scatters" with a black hole, leaving only a slightly more massive black hole, but one with indeterminate global charge as dictated by the no-hair theorem. Heuristically, if one considers "virtual" black hole states of mass M arising from quantum gravity, one can integrate them out to yield global charge violating operators suppressed by powers of M, where M might be as small as M_{Pl}, the Planck mass.

Another indication that gravity might not respect global symmetries comes from wormhole physics. Wormholes are classical solutions to Euclidean gravity that describe changes in topology. Integrating over all wormholes (with a cutoff on their size) yields a low-energy effective action that contains operators of *all* dimensions that violate global symmetries. The natural scale of violation in this case is the wormhole scale, usually thought to be very near (within an order of magnitude or so) M_{Pl}.

Without explicit calculations of these effects, we are left with the following prescription: Due to our lack of understanding of physics at the Planck scale, we have no choice but to interpret theories that do not include gravity in a quantum mechanically consistent way as *effective* field theories with a cutoff at M_{Pl}. If we adhere rigorously to this principle, we are then required to add all higher dimension operators (suppressed by powers of M_{Pl}) consistent with the symmetries of the full theory at M_{Pl}. As discussed above, it seems very unlikely that the full theory respects global symmetries. We note that it would be particularly surprising if the entire theory respects $U(1)_{\rm PQ}$, since this symmetry is already explicitly broken by instanton effects. We should note that similar ideas were noted briefly in the prescient papers of Georgi, Hall, and Wise, and Lazarides, Panagiotakopoulos, and Shafi, however, we are now in a position to be somewhat more specific about the nature of the Planck scale effects in question and to explore their consequences.

We consider first the implications for the axion model. To be specific, we consider a generic invisible axion model in which an electroweak singlet ϕ, charged under $U(1)_{\rm PQ}$, is responsible for spontaneous breaking of the PQ symmetry. We may parametrize ϕ on the vacuum manifold as $\phi = (f_a/\sqrt{2})\ \exp(ia/f_a)$, where a is the axion field. The effects of the

QCD anomaly are to generate a mass for the axion of order $m_a \sim \Lambda^2/f_a$, where Λ is the QCD scale. A variety of astrophysical and cosmological constraints on the axion force f_a into a narrow range of $10^9\mathrm{GeV} \lesssim f_a \lesssim 10^{12}\mathrm{GeV}$ for standard axions, or in a still narrower range around $10^7\mathrm{GeV}$ for hadronic axions.

The instanton induced potential for a takes the form:

$$V(a) = \Lambda^4 \cos(a/f_a + \theta). \tag{35}$$

where θ is the QCD theta angle in a basis where the quark mass matrix is real. While dominating the path integral with instantons is probably a bad approximation in an unbroken gauge theory like QCD, there are rigorous results showing that the minimum of $V(a)$ occurs at strong-CP conserving values.

One possibility is that gravity does not respect $U(1)_{\mathrm{PQ}}$ at all, as is the case if wormhole effects are large. In this case, one should include renormalizable operators such as

$$\Delta V(\phi) \sim M_W^2 \phi^2 + \mathrm{h.c.} \tag{36}$$

Here M_W is the wormhole scale, which is expected to be of the order of the Planck mass. With the addition of these operators, the PQ symmetry is strongly broken and axions never arise at all.

A second possibility is the $U(1)_{\mathrm{PQ}}$ is only broken through non-renormalizable operators of higher dimension. This can occur if either wormhole effects are suppressed or if the PQ symmetry is automatic, i.e., it is present "automatically" when one includes all renormalizable terms consistent with a given gauge group. As we shall see below, higher dimension operators will also spoil the axion solution to the strong-CP problem except possibly in the case of an automatic PQ symmetry, where gauge symmetries can eliminate operators up to some required high dimension.

We now explore the effect upon the axion potential of dimension D operators such as

$$O_D = \frac{\alpha_D}{{M_{Pl}}^{D-4}}\, \phi^{*a}\phi^b + \mathrm{h.c.} \qquad (a \neq b;\;\; a+b = D), \tag{37}$$

which explicitly break $U(1)_{PQ}$. Operators of dimension D will modify the axion potential of Eq. (35):

$$V(a) = \Lambda^4 \cos(a/f_a + \theta) + \sum \Delta_n \cos(na/f_a + \delta_n) \qquad (n = D,\ D-2,\ D-4, \ldots), \tag{38}$$

where $\Delta_n \sim \alpha_D f_a^D/{M_{Pl}}^{D-4}$, and δ_n is a phase angle. Let us simply analyze the $n = 1$ contribution. The extra contribution will shift the minimum of the axion potential away from the strong-CP conserving minimum of $\langle a/f_a \rangle = -\theta$. Unless $\epsilon = \langle a/f_a \rangle + \theta$ is less than 10^{-9} the amount of CP violation obtained will be in conflict with experiment. The minimum of the axion potential is now determined by $f_a V'(a) \simeq \Lambda^4\epsilon + \Delta_1 \sin(\epsilon - \theta + \delta_1) = 0$. The magnitude of $\sin(\epsilon - \theta + \delta_1)$ will not, in general, be small, and $\epsilon \sim \Delta_1/\Lambda^4$.

Since we know $\epsilon < 10^{-9}$, $\Delta_1 < 10^{-9}\Lambda^4$. For dimension D operators, we expect $\Delta_1 \sim \alpha_D f_a^D/{M_{Pl}}^{D-4}$. Using $\Lambda = 10^{-1}\mathrm{GeV}$, the limit on ϵ translates into the following limit on the dimension D of the operator as a function of f_a and α_D:

$$D \lesssim \frac{89 + \log \alpha_D}{9 - \log(f_a/10^{10}\mathrm{GeV})}. \tag{39}$$

If Eq. (39) is satisfied, it is very simple to show that the higher-dimension operators will have an insignificant effect on the axion mass. In fact, the zero temperature axion mass is just $m_a \sim \Lambda^2(1+\epsilon)/f$. However, we should note that the temperature dependence of the axion mass is quite different in the presence of higher dimensional operators. In particular, the mass induced by the higher dimension operators is *always* "turned on." This may affect axion cosmology in interesting ways.

These results at first seem puzzling, since low-energy physics is not in general sensitive to physics at the Planck scale. However, Nambu–Goldstone bosons have the peculiar property that although they are massless (or very light in the case of pseudo-Nambu–Goldstone bosons such as the axion), they are not, properly speaking, part of the low-energy theory as evidenced by the fact that self-couplings, and couplings to light fields are suppressed by a power of a large mass scale. The fact that a light particle such as the axion is part of the high-energy sector accounts for its interesting properties, but also renders it susceptible to high-energy corrections.

In a generic invisible-axion model, there is no reason why a term such as ϕ^5/M_{Pl} could not be generated (here ϕ is a gauge-singlet field). This term would give rise to unacceptable shifts in $\overline{\theta}$ unless $\alpha_D \lesssim 10^{-44-\log(f_a/10^{10}\mathrm{GeV})}$, which is remarkably small. Is there any to avoid this problem?

There are, in fact, ways to construct axion models which suppress higher dimensional operators as needed. This construction is based on the notion of automatic PQ symmetries, as described above. We first consider a supersymmetric automatic model based on the gauge group $E_6 \times U(1)_X$. The superfield content of the model is some number of $\mathbf{27}$'s with X charges ± 1 and a $\overline{\mathbf{351}}$ with X charge 0. The most general renormalizable, gauge-invariant superpotential will only contain terms of the form $\mathbf{27}_1 \cdot \mathbf{27}_{-1} \cdot \overline{\mathbf{351}}_0$, where the subscripts denote the $U(1)_X$ charges. This automatically gives rise to a PQ symmetry in which the $\mathbf{27}$'s have PQ charge $+1$ and the $\overline{\mathbf{351}}$ has PQ charge -2. The lowest dimension operators consistent with gauge invariance in the superpotential that break the PQ symmetry are terms like $\mathbf{27}^6$, $\overline{\mathbf{351}}^6$, and $\mathbf{27}_1 \cdot \mathbf{27}_{-1} \cdot (\overline{\mathbf{351}}_0)^4$. These will then give rise to dimension 10 operators in the effective Lagrangian. Furthermore, it is relatively easy to see that we can break the gauge symmetries and the PQ symmetry spontaneously in such a way so that the final PQ symmetry (a linear combination of the original PQ symmetry and some broken gauge symmetries) is broken around 10^{10} GeV.

It is also possible to construct automatic PQ models based on supersymmetric $SU(N)$ GUT's that suppress higher dimension operators to any desired level for sufficiently large N. Models of this type without exotic fermions must all have at least four different chiral matter irreducible representations whose Young tableaux consist of a single column. Needless to say, these are exceedingly unattractive models. They will tend to have many extra families, which in addition to a host of phenomenological problems, will possibly destroy the asymptotic freedom of QCD.

Planck scale physics may also significantly affect the other solutions for the strong-CP problem. As described above, the second class of solutions are based upon models where CP is softly or spontaneously broken. How they fare under Planck scale physics depends on whether dimension four operators are generated, or whether only higher dimension operators appear. If renormalizable operators can be generated, then the violation of CP by Planck scale effects will give rise to a $\mathrm{tr}\, G\widetilde{G}$ term, thus regenerating the strong-CP problem (we should note, however, that the coefficient of such a term could be exponentially suppressed if it appeared in some controlled semiclassical expansion about some classical configuration).

Let us next consider the case in which only non-renormalizable operators are generated by Planckian physics. In this case, all models with fields that acquire vacuum expectation values well below the Planck scale (typically the weak scale), will generate corrections to $\overline{\theta}$ that are highly suppressed by powers of M_{Pl}. In essence, this is nothing more than a restatement of the effective field theory philosophy: as long as we consider physics at energies below the cutoff of our theory, the dominant effects come from the renormalizable operators in the theory. This way of thinking about effective field theories explains why the PQ solution is so susceptible to possible effects of gravity. The problem is that the PQ scale is too close to

M_{Pl} while the constraints on $\bar{\theta}$ are too tight.

Although we have seen that wormholes are troublesome for models that claim to solve the strong-CP problem, there is some indication that wormhole effects themselves might drive the QCD $\bar{\theta}$ parameter to a CP conserving value. Within the framework of Coleman's wormhole calculus (which has since been shown to be naive in some respects), $\bar{\theta}$ became a function of the wormhole parameters. The implementation of Coleman's prescription for determining the value of these parameters was then shown to set $\bar{\theta}$ to a CP conserving value. It is not impossible that a more sophisticated approach to the wormhole calculus would still lead to a similar situation. However, until a better understanding of wormholes and quantum gravity in general is reached, this will remain a conjecture.

We see that Planck-scale physics can have dramatic effects on axion physics. If one wants to pursue the axion solution to the strong-CP problem, automatic models such as those presented here are probably the only consistent approach that can be taken. We have also argued that the other known solutions are essentially unaffected by gravity. The essential difference between the PQ and the non-axionic solutions is due to the sensitivity of the Nambu–Goldstone boson to physics at energies near the scale of spontaneous symmetry breaking. It remains to be seen whether other facets of the axion scenario, such as the axion energy density crisis will be modified by the effects considered here. For more details, the see the recent papers of Holman *et al.* (1992b), Kamionkowski *et al.*, (1992b), and Barr *et al.* (1992).

In conclusion, any model which depends on the dynamics of Nambu–Goldstone modes will be extremely sensitive to physics at very high energies. Texture can by no means be considered a robust prediction of unified theories. This is most discouraging for the texture scenario. On the other hand, if texture is discovered, then this will have profound implications not only for theories of structure formation, but for Planck-scale physics. The same is true for axions. A cosmological discovery of either texture or axions would tell us something about particle physics at the highest energies. What better way to close lectures on the implications of cosmology for particle physics.

Finally there are other creatures that might be produced in cosmological phase transitions. Non-topological solitons, or Q-balls, (Frieman *et al.*, 1988), and electroweak strings (Achucarro and Vachaspati, 1991).

The lesson for cosmological phase transitions is that even with unlimited energy, accelerators are the wrong tool to probe the non-perturbative sector of field theories. Early-Universe phase transitions continue to provide the best arena for the study of aspects of particle-physics theories related to coherent, soliton-like objects. The only plausible site for the production of objects such as monopoles, strings, walls, sphalerons, and the like is an early-Universe phase transition. All of these can have very significant implications for the evolution of the Universe. Shpaleons, as well as other solitons produced in the electroweak transition, have some promise of a cosmological payoff. Of course there is an enormous difference between finding a soliton-like solution to the field equations and finding solitons in the Universe. However, even if they are not are found, the techniques developed for their study will be useful additions to the theorist's toolbox.

4. Acknowledgements

This work was supported in part by the Department of Energy and the National Aerounatics and Space Administration (Grant NAGW–2381) at at Fermilab. I am happy to acknowledge collaborations with Marcelo Gleiser, Rich Holman, Steve Hsu, Tom Kephart, Rick Watkins, and Larry Widrow in the work reported here.

5. References

1. Achucarro A. and Vachaspati T., 1991, *Phys. Rev. D* **44**, 3067.
2. Anderson G. and Hall L. J., 1992, *Phys. Rev. D* **145**, 2685.
3. Barr S. and Seckel D, 1992, *Phys Rev D* **46**, 539.
4. Cohen A., Kaplan D., and Nelson A., *Ann. Rev. Nucl. Part. Sci.* **43**, 27.
5. De Gennes P. G., *The Physics of Liquid Crystals*, (Clarendon, Oxford, 1975).
6. Dine M., Leigh R., Huet P., Linde A., and Linde D., 1992, *Phys. Rev. D* **46**, 550.
7. Dolan L. and Jackiw R., 1974, *Phys. Rev. D* **9**, 3320.
8. Fernández L. A., Ruiz-Lorenzo J. J., Lombardo M., and Tarancón A., 1992, *Phys. Lett.* **B277**, 485.
9. Frieman J. A., Gelmini G. B., Gleiser M., and Kolb E. W., 1988, *Phys. Rev. Lett.* **60**, 2101.
10. Ginsparg P., 1980, *Nucl. Phys.* **B170**, 388.
11. Gleiser M., Kolb E. W., and Watkins R., 1991, *Nucl. Phys.* **B364**, 411.
12. Gleiser M. and Kolb E. W., 1993, *Phys. Rev. Lett.* **69**, 1304.
13. Holman R., Hsu S. H., Kolb E. W., Watkins R., and Widrow L. M., 1992a, *Phys. Rev. Lett.* **69**, 1489.
14. Holman R., Hsu S. H., Kephart, T. W., Kolb E. W., Watkins R., and Widrow L. M., 1992b, *Phys. Lett.* **B282**, 132.
15. Kamionkowski M. and March-Russell J., 1992a, *Phys. Rev. Lett.* **69**, 1485.
16. Kamionkowski M. and March-Russell J., 1992b, *Phys. Lett.* **B282**, 137.
17. Kuzmin V. A., Rubakov V. A., and Shaposhnikov M. E., 1985, *Phys. Lett.* **155B**, 36.
18. Pen U.-L., Spergel D. N., and Turok N., 1993, *Cosmic structure formation and microwave anisotropies from global field ordering*, Princeton preprint.
19. Weinberg S., 1974 *Phys. Rev. D* **9**, 3357.